

HT1220 Report

RECORD INFORMATION

Gene ID: 1220
Sequence ID: 1220
Protein ID: 1220
Sequence name: thrombospondin 1, alt. transcript 1
Genome: nucleus
Taxon: Homo sapiens
Locus: 1220
Common Name: thrombospondin 1
Role ID: 40

Coding sequence length: 3513 nt
Transcript sequence length: 5722 nt
Expression data: THC201673

ACCESSION DATA

HT1220 is derived from accessions(s):

SP:P07996 (THROMBOSPONDIN 1 PRECURSOR.)
GB:X04665 (Human mRNA for thrombospondin)
GB:X14787 (Human mRNA for thrombospondin)
GB:U12471 (thrombospondin-p50 {Homo sapiens})
GB:M99425 (Human thrombospondin mRNA, 3' end.)
PIR:G01478 (thrombospondin-p50 - human (fragment))
GB:U12471 (Human thrombospondin-1 gene, partial cds.)
GB:J04835 (Human thrombospondin gene, exons 1, 2 and 3.)
GB:M25631 (Homo sapiens (clone lambda-TS-33) thrombospondin (THBS) mRNA, 5' end.)

ALTERNATIVE SPLICE INFORMATION

Alternative splice forms for this gene:

HT3987 thrombospondin 1, alt. transcript 2

MAPPING DATA

GDB accession(s) for this gene:

GDB ID: Symbol

FIGURE 1A

cDNA FEATURES

Feature	End 5	End 3
coding_seq	112	3624
3'UT	3625	5722
spjunc_h	1235	1236

SEQUENCE

nucleotide:

ggacgcacaggaattccccgcgccccctccagccctcgccgcctcgccaccgctcccggc
 cgccgcgctccgggtacacacaggatccctgctgggcaccaacagctccaccatggggctg
 gcctggggactagggctcctgttccctgatgcatgtgtgtggcaccaaccgcatccagag
 tctggcggagacaacagcggtttgacatccttgaactcacggggccgcccgcgaagggg
 tctggggcgccgactgggtgaagggccccgaccctccagcccagctttccgcatcgaggat
 gccaacctgatccccctgtgcctgatgacaagtccaagacctgggtggatgctgtgcgg
 gcagaaaagggtttcctccttctggcatccctgaggcagatgaagaagaccggggcacg
 ctgctggccctggagcggaaaagaccactctggccaggctcttcagcgtgggtgtccaatggc
 aaggcggggcacccctggacctcagcctgaccgtccaaggaaagcagcacgtgggtgtctgtg
 gaagaagctctcctggcaaccggccagtggaagagcatcacctgtttgtgcaggaagac
 agggcccagctgtacatcgactgtgaaaagatggagaatgctgagttggacgtccccatc
 caaagcgtcttcaccagagacctggccagcatcgccagactccgcatcgcaaaggggggc
 gtcaatgacaatttcagggggtgtctgcagaatgtgaggtttgtctttggaaccacacca
 gaagacatctcaggaacaaggctgtccagctctaccagtgctcctcctcaccttgac
 aacaacgtgggtgaatgggtccagccctgccatccgcactaacctacattggccacaagaca
 aaggacttgcaagccatctgcggcatctcctgtgatgagctgtccagcatgggtcctggaa
 ctcaggggctgcgcaccattgtgaccacgctgcaggacagcatccgcaaagtgaactgaa
 gagaacaaagagtggcccaatgagctgaggcggcctcccctatgctatcacacggagtt
 cagtacagaaataacgaggaatggactgttgatagctgcactgagtgctcactgtcagaac
 tcagttaccatctgcaaaaagggtgtcctgccccatcatgcctgctccaatgccacagtt
 cctgatgggaatgtgtcctcgtgttggtggccagcagctctgcggacgatgggtgggtct
 ccattggtccgagtggaacctcctgttctacgagctgtggcaatgggaattcagcagcgcgcc
 cgctcctgcgatagcctcaacaaccgatgtgagggctcctcggtccagacacggacctgc
 cacattcaggagtggtgacaaaagatttaaacaggatgggtggctggagccactgggtccccg
 tggatcatcttgttctgtgacatgtgggtgatgggtgtgatcacaaggatccggctctgcac
 tctcccagccccagatgaatgggaaaccctgtgaaggcgaagcgcgggagaccaaagcc
 tgcaagaagacgcctgccccatcaatggaggctggggctcctgggtcaccatgggacatc
 tgttctgtcaccctgtggaggagggtacagaaacgtagtcgtctctgcaacaaccccgca
 cccagtttggaggcaaggactgcgttgggtgatgtaacagaaaaccagatctgcaacaag
 caggactgtccaattgatggatgcctgtccaatccctgctttgocggcgtgaagtgtact
 agctaccctgatggcagctggaaaatgtgggtgcttgtccccctgggttacagtggaaaatggc
 atccagtgcacagatgttgatgagtgcaagaagtgcctgatgcctgcttcaaccacaat
 ggagagcaccgggtgtgagaacacggaccceggctacaactgcctgcccctgccccccacgc
 ttacccggtcacagcccttcgggccagggtgtcgaaatgccacgggccaacaaacagggtg
 tgcangccccgtaacccctgcaaggatgggacccacgactgcaacaagaacgccaagtgc
 aactacctgggcccactatagcgacccccatgtaccgctgcagtgcaagcctggctacgct
 ggcaatggcatcatctgcggggaggacacagacctggatgggtggcccaatgagaacctg
 gtgtgcgtggccaatgcgacttaccactgcaaaaaggataattgccccaaccttcccaac
 tcagggcaggaagactatgacaaggatggaattgggtgatgcctgtgatgatgacgatgac
 aatgataaaattccagatgacaggggacaactgtccattccattacaaccagctcagtat
 gactatgacagagatgatgtgggagaccgctgtgacaactgtccctacaaccacaacca

FIGURE 1B

gatcaggcagacacagacaacaatggggaaggagacgcctgtgctgcagacattgatgga
gacgggtatcctcaatgaacgggacaactgccagtagctctacaatgtggaccagagagac
actgatatggatgggggttggagatcagtgtagacaattgcccccttggaaacacaatccggat
cagctggactctgactcagaccgcattggagatacctgtgacaaacaatcaggatattgat
gaagatggccaccagaacaatctggacaactgtccctatgtgcccattgccaaaccaggct
gaccatgacaagaatggcaaggagatgcctgtgaccacgatgatgacaacgatggcatt
cctgatgacaaggacaactgcagactcgtgccccaatcccgaccagaaggactctgacgyc
gatggctcgaggatgcctgcaaagatgattttgaccatgacagtggtgccagacatcgat
gacatctgtcctgagaatgttgacatcagtgagaccgatttccgcgcgattccagatgatt
cctctggacccccaaaggacatccccaaatgaccctaactgggttgtagcccatcagggt
aaagaactcgtccagactgtcaactgtgatcctggactcgtgttaggttatgatgagttt
aatgctgtggacttcagtgggaccttcttcatcaacaccgaaaggagcagatgactatgct
ggatttgtcttgggtaccagtcacgcgcgcttttatgttgtagtggaagcaagtc
acccagtcctactgggacaccaacccccacgagggtcagggtatactcgggcctttctgtg
aaagtgttaactccaccacagggcctggcgagcacctgcggaacgcctgtggcacaca
ggaaacacccctggccagggtgcgcacccctgtggcatgaccctcgtcacataggctggaa
gatttcacgcctacagatggcgtctcagccacaggccaaagacgggtttcattagagtg
gtgatgtatgaagggaagaaaatcatggctgactcaggaccatctatgataaaacctat
gctgggtgtagactagggttgtttgtcttctctcaagaaatgggtgttcttctctgacctg
aaatagcaatgtagagatccctaactcatcaaattgttgattgaaagactgatcataaacc
aatgctgggtattgcaccttcttggaaactatgggcttgagaaaacccccaggatcacttctc
cttggttctcttcttcttctgtgcttgcatcagtggtgactcctagaacgtgacgacctgac
ctaaagaaaatgcagtttctcaaaaacagactcatcagcattcagcctccaatgaataagac
atcttccaagcatataaacaattgcttttggtttctcttttgaaaaagcatctacttgcttc
agttgggaagggtgccatttccactctgcctttgtcacagagcagggtgctattgtgaggc
catctctgagcagtggtgactcaaaagcattttcaggcatgtcagagaaggaggactcact
agaattagcaaacaaaaccacctgacatcctccttcagggaacacggggagcagaggcca
aagcactaaggggaggggcgcatacccgagacgattgtatgaagaaaatattggaggaaactg
ttacatgttccgtactaagtcattttcagggggattgaaagactattgctggatttcatga
tgctgactggcggttagctgatttaacccatgtaaataggcacttaaatagaagcaggaaag
ggagacaaagactggcttcttggacttctcctcctgatccccaccttactcatcacttgc
ctggteacattgaaattgggtggcttcatcttagatgtagcttggtgcagatgtagcaggaa
aataggaaaaacctaccatctcagtgagcaccagctgcctcccaaaggaggggcagccgtg
cttatattttttatgggttacaatggcacaaaatttatatcaacotaactaaaacatttctt
ttctcttttttccgtaattactaggtagttttctaattctctcttttggaaagtatgattt
ttttaaagtctttacgatgtaaaaatatttattttttacttattctggaagatctggctga
aggattattcatggaacaggaagaagcgtaaagactatccatgtcatctttgttgagagt
cttcgtgactgtaagattgttaaatacagattattttattaactctgttctgctggaaatt
taggcttcatacgggaagtggttgagagcaagtagttgacatttatcagcaaatctcttg
caagaacagcacaaaggaaaatcagtcataaagctgctctgcccccttggtgctcagagtgg
atgttatgggatttcttttttctctgttttatcttttcaagtggaaattagttgggtatcc
atttgcaaatgttttaaatgcaaaagaaagccatgaggtcttcaatactgttttaccacca
tcccttggtgcatatttccaggggagaaggaaagcatatacacttttttcttcttcttcc
aaaagagaaaaaaatgacaaaagggtgaaacttacatacaaatattacctcatttgttggtg
tgactgagtaagaaattttggatcaagcggaaagagtttaagtgtctaaacaaacttaaa
gctactgtagtacctaaaagtcagtggttgtagatagcataaaaactctgacagagaagta
ttcccaataaggaaatagcattgaaatgttaaatataaattttctgaaagtattgtttttt
tctatcatctgggtataccattgctttatttttataaattattttctcattgccattggaa
tagaatattcagattgtgtagatatgctattttaataatttatcaggaaatactgctgtg
agagtttagtattttctatttttatataatgtttgcacactgaattgaagaattgttggttt
tttcttttttttggtttttttttttttttttttttttggcttttgacctcccattttta
ctatttgccaatacctttttctaggaatgtgcttttttttgtacacatttttatccattt
cacattctaaagcagtgtaagtgtagattactgtttcttatgtacaaggaacaacaata
aatcatatggaaatttatattt

protein:

MGLAWGLGVLFLMHVCGTNRI PESGGDNSVFDI FELTGAARKGSORRLVKGPDPSSPAFR

FIGURE 1C

IEDANLIPPVPDDKFQDLVDAVRAEKGFLLLASLRQMKTRGTLALERKDHSQQVFSVV
SNGKAGTLOLSLTVQGKHVVSVEEALLATGQWKSITLQVQEDRAQLYIDCEKMEAELO
VPIQSVFTRDLASIARLRIAAGGVNDNFQGVLOQNVRFVFGTTPEDILRNKGCSSSTSVLL
TLONNVVNGSSPAIRTNVYIGHKTKDLQAIQGISCELSMVLRLGLRTIVTTLQDSIRK
VTBENKELANELRRPPLCYHNOVQYRNNEBWTVDSCTECHCQNSVTICKKVSCPINPCSN
ATVPDGECCPRCWPSDSADGWSPWSEWTSCSTSCGNGIQQRGRSCDSLNNRCEGSSVQT
RTCHIQECDKRFKQDGGWSHSPWSSCSVTGCGDVITRIRLNCNSPSPQMNGKPCEGEARE
TKACKKDACPINGGWGPWSPWDICSVTCGGGVQKRSRLCNPAPQFGGKDCVGDVTENQI
CNKQDCPIDGCLSNPCFAGVKCTSYPDGSKWCGACPPGYSGNGIQCTDVDECKEVPDACF
NHNGEHRCENTDPGYNCLPCPPRFTGSQPFQGVQVEHATANKQVCKPRNPCTDGTHDCNKN
AKCNYLGHYSDPMYRCECKPGYAGNGIICGEDTDLGWPENLVCVANATYHCKKDNCPN
LPNSGQEDYDKDGIQDACDDDDNDKI PDDRDNCPFHYNPAQYDYDRDDVGDRCDNCPYN
HNPDAQADTDNNGEGDACAADIDGDGILNERDNCQYVYNVDQRDTDMGVDGQCDNCPLEH
NPDQLSDSDRIGDTCNNQDIDEDGHQNNLDNCPYVPANQADHDKGKGDACDHDDDN
DGIPDDKDNCRLLVPNPDQKSDGDGRGDACKDDFDHDSVPDIDDICPENVDISETDFRRF
QMIPLDPKGT SQNDPNWVVRHQGKELVQTVNCDPGLAVGYDEFNAVDFSGTFFINTERDD
DYAGFVFGYQSSSRFYVVMWKQVTQSYWDTNPTRAQGYSGLSVKVNSTTGPEHLRNAL
WHTGNTPGQVRTLWHDPRHIGWKDFTAYRWRLSHRPKTGFIRVVMYEGKKIMADSGPIYD
KTYAGGRLGLFVFSQEMVFFSDLXYECRDP

FIGURE 1D

SEQUENCE- TSP-2

nucleotide:

acggcatcca	gtacagaggg	gctggacttg	gacccctgca	gcagccctgc
acaggagaag	cggcatataa	agccgcgctg	cccgggagcc	gctcggccac
gtccaccgga	gcatcctgca	ctgcagggcc	ggtctctcgc	tccagcagag
cctgcgcctt	tctgactcgg	tccggaacac	tgaaccaggt	catcactgca
tcttttttggc	aaaccaggag	ctcagctgca	ggaggcagga	tggctctggag
gctggtcctg	ctggctctgt	gggtgtggcc	cagcacgcaa	gctggtcacc
aggacaaaga	cacgaccttc	gaccttttca	gtatcagcaa	catcaaccgc
aagaccattg	gcgccaagca	gttccgcggg	cccgaacccg	gcgtgccggc
ttaccgcttc	gtgcgctttg	actacatccc	accggtgaac	gcagatgacc
tcagcaagat	caccaagatc	atgcggcaga	aggagggtt	cttcctcacg
gcccagctca	agcaggacgg	caagtccagg	ggcacgctgt	tggctctgga
gggccccggg	ctctcccaga	ggcagttcga	gatcgtctcc	aacggccccg
cggacacgct	ggatctcacc	tactggattg	acggcaccgg	gcattgtggc
tccctggagg	acgtcggcct	ggctgactcg	cagtgggaaga	acgtcaccgt
gcaggtggct	ggcgagacct	acagcttgca	cgtgggctgc	gacctcatag
gaccagttgc	tctggacgag	cccttctacg	agcacctgca	ggcggaaaag
agccggatgt	acgtggccaa	aggctctgcc	agagagagtc	acttcagggg
tttgcttcag	aacgtccacc	tagtgtttga	aaactctgtg	gaagatattc
taagcaagaa	gggttgccag	caaggccagg	gagctgagat	caacgccatc
agtgagaaca	cagagacgct	gcgcctgggt	ccgcatgtca	ccaccgagta
cgtgggcccc	agctcggaga	ggaggcccga	ggtgtgcgaa	cgctcgtgcg
aggagctggg	aaacatggtc	caggagctct	cggggctcca	cgctcctcgtg
aaccagctca	gcgagaacct	caagagagtg	tcgaatgata	accagtttct
ctgggagctc	attggtggcc	ctcctaagac	aaggaacatg	tcagcttgct
ggcaggatgg	ccggttcttt	gcggaaaatg	aaacgtgggt	ggtggacagc
tgcaccacgt	gtacctgcaa	gaaatttaaa	accatttgcc	accaaatac
ctgcccgcct	gcaacctgcg	ccagtccatc	ctttgtggaa	ggcgaatgct
gcccttcctg	cctccactcg	gtggacgggtg	aggagggtcg	gtctccgtgg
gcagagtggg	cccagtgtct	cgtgacgtgt	ggctctggga	cccagcagag
aggccgggtc	tgtgacgtca	ccagcaacac	ctgcttgggg	ccctcgatcc
agacacgggc	ttgcagtctg	agcaagtgtg	acaccgcgat	ccggcaggac
ggcggctgga	gccactggtc	accttgggtc	tcatgctctg	tgacctgtgg
agttggcaat	atcacacgca	tccgtctctg	caactcccca	gtgcccagga
tggggggcaa	gaattgcaaa	gggagtggcc	gggagaccaa	agcctgccag
ggcgccccat	gcccgaatcg	tggccgctgg	agcccctggg	cccgtgggtc
ggcctgcact	gtcacctgtg	ccggtgggat	ccgggagcgc	acccgggtct
gcaacagccc	tgagcctcag	tacggaggga	aggcctgcgt	gggggatgtg
caggagcgtc	agatgtgcaa	caagaggagc	tgccccgtgg	atggctgttt
atccaacccc	tgcttcccgg	gagcccagtg	cagcagcttc	cccgatgggt
cctggctcatg	cggcttctgc	cctgtgggct	tcttgggcaa	tggcaccac
tgtgaggacc	tggacgagtg	tgccctgggtc	cccgacatct	gcttctccac
cagcaagggtg	cctcgctgtg	tcaacactca	gcctggcttc	cactgcctgc
cctgcccgcc	ccgatacaga	gggaaccagc	ccgtcggggg	cggcctggaa
gcagccaaga	cggaaaagca	agtgtgtgag	cccgaacc	catgcaagga
caagacacac	aactgccaca	agcacgcgga	gtgcatctac	ctgggtcact
tcagcgaccc	catgtacaag	tgcgagtgcc	agacaggcta	cgcgggcgac
gggctcatct	gcggggagga	ctcggacctg	gacggctggc	ccaacctcaa
tctggtctgc	gccaccaacg	ccacctacca	ctgcatcaag	gataactgcc

FIGURE 2A

atgaggagtt tatgatggaa ccttaatatata taatgttgcc agcgatttta
 gttcaatatt tggtactggt atctatctgc tgtatatgga attcttttaa
 ttcaaacgct gaaaacgaat cagcatttag tcttgccagg cacaccaat
 aatcagtcac gtgtaatatg cacaagtttg tttttgtttt tgttttttt
 gttggttggt ttttttgctt taagttgcat gatctttctg caggaaatag
 tcaactcatcc cactccacat aaggggttta gtaagagaag tctgtctgtc
 tgatgatgga tagggggcaa atctttttcc cctttctggt aatagtcac
 acatttctat gccaaacagg aacgatccat aacttttagtc ttaatgtaca
 cattgcattt tgataaaaatt aattttgttg tttcctttga ggttgatcgt
 tgtgttggtt tgctgcactt tttacttttt tgcgtgtgga gctgtattcc
 cgagacaacg aagcgttggtg atacttcatt aaatgtagcg actgtcaaca
 gcgtgcagggt tttctgtttc tgtgttggtg ggtcaaccgt acaatggtgt
 gggaatgacg atgatgtgaa tatttagaat gtaccatatt ttttgtaa
 tatttatggt tttctaaaca aatttatcgt ataggttgat gaaacgtcat
 gtgttttgcc aaagactgta aatatttatt tatgtgttca catggtcaaa
 atttcaccac tgaaaccctg cacttagcta gaacctcatt tttaaagatt
 aacaacagga aataaattgt aaaaaagggt ttct

protein:

MVWRLVLLALWVWPSTQAGHQDKDTTFDLFSISNINRKTIGAKQFRGPDGPVPAYRF
 VRFDYIPPVNADDLSKITKIMRQKEGFFLTAQLKQDGKSRGTLLEGLSQRQFE
 IVSNGPADTLDLTYWIDGTRHVVSLEDVGLADSQWKNVTVQVAGETYSLHVGCDLIG
 PVALDEPFYEHQAESRMYVAKGSARESHFRGLLQNVHLVFENSVEDILSKKGCQO
 GQGAIEINAISENTETLRLGPHVTTEYVGPSSERRPEVCERSCEELGNMVQELSGLHV
 LVNQLSENLRVSNDNQFLWELIGGPPKTRNMSACWQDGRFFAENETWVVDSTCTCT
 CKKFKTICHQITCPPATCASPSFVEGECCPSCLHSVDGEEGWSWPAEWTQCSVTCGS
 GTQQRGRSCDVTSNTCLGPSIQTRACSLSKCDTRIRQDGGWSHWSPWSSCSVTCGVG
 NITRIRLCNSPVPQMGGKNCKGSGRETKACQGAAPCPIDGRWSPWSPWSACTVTCAGG
 IRERTRVCNSPEPQYGGKACVGDVQERQMCNKRSCPVDGCLSNPCFPGAQCSSFPDG
 SWSCGFPCPVGFLGNGTHCEDLDECALVPDICEFSTSKVPRCVNTQPGFHCLPCPPRYR
 GNQPVGVGLEAAKTEKQVCEPENPCKDKTHNCKHAECIYLGHFSDPMYKCECQTGY
 AGDGLICGEDSDLDGWPNNLVCATNATYHCICKDNCPHLPNSGQEDFDKDGIGDADC
 DDDNDGVTDEKDNCQLLENPRQADYDKDEVGDRCDNCPYVHNPAQIDTDNNGEGDA
 CSVDIDGDDVFNERDNCYVYNTDQRDTDGDGVGDHCDNCPVHNPDQTDVDNDLVG
 DQCDNNEDIDDDGHQNNQDNCYISNANQADHHRDGGQDADCDPDDNDGVPDDRNDNC
 RLVFNPDQEDLDGDGRGDICKDDFDNDNIPDIDDVCPENNAISETDFRNFQMVPLDP
 KGTQIDPNWVIRHQGKELVQTANSDPGIAVGDFEFGSVDFSGTFYVNTDRDDDYAG
 FVFGYQSSSRFYVVMWKQVTQTYWEDQPTRAYGYSGVSLKVVNSTTGTGEHLRNALW
 HTGNTPGQVRTLWHDPRNIGWKDYTAYRWHLTHRPKTGYIRVLVHEGKQVMADSGPI
 YDQTYAGGRLGLFVFSQEMVYFSDLKYECRDI

FIGURE 2C

HT2143 Report

RECORD INFORMATION

Gene ID: 2081
Sequence ID: 2143
Protein ID: 2125
Sequence name: thrombospondin 4
Genome: nucleus
Taxon: Homo sapiens
Locus: 2081
Common Name: thrombospondin 4
Role ID: 40

Coding sequence length: 2886 nt
Transcript sequence length: 3074 nt
Expression data: THC168897

ACCESSION DATA

HT2143 is derived from accessions(s):

SP:P35443 (THROMBOSPONDIN 4 PRECURSOR.)
GB:Z19585 (thrombospondin-4 (Homo sapiens))
GB:Z19585 (H.sapiens mRNA for thrombospondin-4)
PIR:A55710 (thrombospondin 4 precursor - human)

cDNA FEATURES

Feature	End 5	End 3
coding_seq	28	2913
3'UT	2914	3074

SEQUENCE

nucleotide:

gaattccggggagcaggaagagccaacatgctggccccgcgcggagccgcgctcctcctg
ctgcacctggtcctgcagcggtagcggcaggcgcccaggccacccccagggtcttt
gaccttctcccatcttccagtcagaggctaaacccaggcgctctgctgccagtcctgaca
gaccccgccctgaatgatctctatgtgatttccaccttcaagctgcagactaaaagttca
gccaccatcttcggtctttactcttcaactgacaacagtaaatattttgaatttactgtg
atgggacgcttaagcaaacccatcctccgttacctgaagaacgatgggaaggtgcatttg

FIGURE 3A

gtgggttttcaacaacctgcagctggcagacggaaggcgccacaggatcctcctgaggctg
agcaatttgcagcgagggggccggctccctagagctctacctggactgcacccagggtggat
tcogttcacaatctccccagggcctttgctggccccctccagaaacctgagaccattgaa
ttgaggactttccagaggaagccacaggacttcttggaaagagctgaagctgggtgggtgaga
ggctcactgttccaggtggccagcctgcaagactgcttctgcagcagagtgagccactg
gctgccacagggcacaggggactttaaccggcagttcttgggtcaaatgacacaattaaac
caactcctgggagaggtgaaggacottctgagacagcagggttaaggaaacatcatttttg
cgaaacaccatagctgaatgccagggttgccgttctctcaagtttcagtctccgacccca
agcacgggtgggtgcggccgggtccccctgcacggcccaacacggccacctcgctgggtgtgac
tccaacccatgtttccgaggtgtccaatgtaccgacagtagagatggcttccagtggtggg
ccctgccccgaggggtacacaggaaacgggatcacctgtattgatgttgatgagtgcaaa
taccatccctgctacccggggtgacactgcataaatttgtctcctggcttcagatgtgac
gctgcccagtggggtctcacaggggcccatgggtgcagggtgttgggatcagttttgccaag
tcaacaagcaggtctgcactgcattgatgagtgctgaaatggagcgtgcgttcccaac
tcgatctgcgttaatactttgggatcttaccgctgtgggccttgtaagccgggggtatact
gggtgatcagataaggggatgcaagtggaagaaactgcagaaacccagagctgaaccct
tgagtgatgaatgccagtgcatggaagagaggcagggggatgtgacatgtgtgtgtggg
gtcgggtggggtggagatggctatatctgtggaaaggatgtggacatcgacagttacccc
gacgaagaactgccatgctctgccagggaactgtaaaaaggacaactgcaaatatgtgcca
aattctggccaagaagatgcagacagagatggcattggcgacgcttgtagcagaggatgct
gacggagatgggatcctgaatgagcaggataactgtgtcctgattcataantgtggacca
aggaacagcgataaagatatctttggggatgcctgtgataactgcctgagtgctttaa
aacgaccagaaaagacacggatggggatgggaagaggagatgcctgtgatgatgacatggat
ggagatgggaataaaaaacattctgggacaactgccccaaaatttcccaatcgtagccaacgg
gacaaggatgggtgatgggtgtgggggatgcctgtgacagttgtcctgatgtcagcaaccct
naccagctctgatgtggataatgatctgggtggggactcctgtgacaccaatcaggacagt
gatggagatgggcaccaggacagcacagacaactgccccacggctattaacagtgcccag
ctggacaccgataaaggatgggaattgggtgacgagtgatgatgatgatgacaatgatgg
atcccagacctgggtgccccctggaccagacaactgcccggctgggtccccaaaccagcccag
gaggatagcaacagcgacggagtgaggagacatctgtgagtgctgactttgaccaggaccag
gtcatcgatcggatcgacgctctgccagagaacgcagagggtcacctgaccgacttcagg
gcttaccagacgggtgggctggatcctgaaagggtatgccagatcgatcccaactgggtg
gtcctgaaccagggtcatggagattgtacagaccatgaacagtgatcctggcctggcagtg
gggtacacagcttttaattggagttgacttcgaagggtaccttccatgtgaataccagaca
gatgatgactatgcaggctttatctttgggtaccaagatagctccagcttctacgtgggtc
atgtggaagcagacggagcagacatattgggaagccaccccatccgagcagttgcagaa
cctggcattcagctcaaggtctgtgaagtctaagacagggtccaggggagcatctccggaac
tccctgtggcacacgggggacaccagtgaccaggtcaggctgctgtggaaggactccagg
aatgtgggttggaaggacaaggtgtcctaccgctgggttccctacagcacaggccccagggtg
ggctacatcagggtacgattttatgaaggctctgagttgggtgggtgactctggcgtcacc
atagacaccacaatgcgtggaggccgacttggcgttttctgcttctctcaagaaaacatc
atctgggtccaaacctcaagtatcgctgcaatgacaccatccctgaggacttccaagagttt
caaaccagaaatttcgaccgcttcgataattaaaccaagggaagcaatctgtaactgcttt
tcggaacactaaaaccatataatatttaacttcaattttcttttagcttttaccaccccaa
atatatcaaaaacgttttatgtgaaatgtggcaataaaggagaagagatcatttttataaaaa
aaaaaaaaaaaaaa

protein:

MLAPRGAAVLLLHLVLQRWLAAGAQA TPQVFDLLPSSSQRLNPGALLPVLTDPALNOLYV
ISTFKLQTKSSATIFGLYSSTDNSKYFEBFTVMGRLSKAILRYLKNDGKVHLVVFNNLQLA
DGRHRIRLLRLSNLQRGAGSLELYLDCIQVDSVHNLPRAFAGPSQKPEITELRTFQRKPQ
DFLEELKLVVRGSLFQVASLQDCFLQQSEPLAATGTGDFNRQFLGQMTQLNQLLGVEVKDL
LRQVKETSFLRNTIAECQACGPKLQFSPTSTVVPAPAPAPPTPPRRCDSPNCFRGVQ
CTDSRDGFQCGPCPEGYTGNIGITCIDVDECKYHPCYPGVHCINLSPGFRCDACFVGTGP
MVQGVGISFAKSNKQVCTDIDECRNGACVPNSICVNTLGSYRCGPCKPGYTGDQIRGCKV
ERNCRNPENLPCSVNAQCIERQGDVTCVCGVGWAGDGYICGKDVIDSYDPDEELPC SAR
NCKKDNCKYVPNSGQEDADRDGIGDADDEDADGDGILNEQONCVLIHNVQDRNSDKDIFG
DACDNCLSVLNNQDKD TDGDRGDACDDMDGDGIKNILDNCKFPNRDQRDKDGDGVGD

FIGURE 3B

[illegible][illegible]

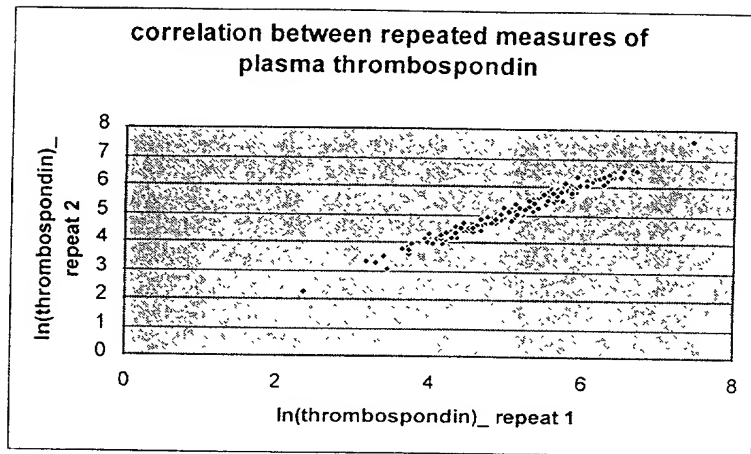


FIGURE 4